were considered. One is discussed in the text (denoted by TSC2, where HWDTT is used), and the second one only uses cases to detect departure from HWE (denoted by TSC1).

The results from the simulations are reported in Table S1. The results show that TSC1 is usually more powerful than TSC2. Note that TSC1 is more powerful than the optimal trend test under the REC model when MAF is small to moderate. But TSC1 is much less powerful than the optimal trend test under the ADD and MUL models. This is because testing HWE has little power under these two models. TSC1 catches some power under the DOM model, but it is slightly less powerful than the optimal-trend test. On the other hand, when the genetic model is unknown, we cannot use the optimal-trend test. However, we compare the TSC1 with the robust test MAX3, which does not require that we know the genetic model. Table S1 shows that, except for the REC model, MAX3 is more powerful than TSC1.

Yong Zang,[1] Wing K. Fung,[1] and Gang Zheng[2,*]
[1]Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China; [2]Office of Biostatistics Research, Division of Prevention and Population Sciences, National Heart, Lung, and Blood Institute, Bethesda, MD 20892-7913, USA
*Correspondence: zhengg@nhlbi.nih.gov

Supplemental Data

Supplemental Data include two figures and one table and are available with this article online at http://www.ajhg.org/.

Acknowledgments

We would like to thank Yaning Yang for some helpful discussions that brought our attention to the tail strength. The work of Y. Zang and W.K. Fung were partially supported by The Croucher Foundation and China Natural Science Foundation (no. 10701067).

Web Resources

The URL for data presented herein is as follows:

The R program (TSC.txt) used in the simulation can be downloaded from the website: www.statisticalsource.com/software/TSC1.txt.

References

1. Wang, J., and Shete, S. (2008). A test for genetic association that incorporates information about deviation from Hardy-Weinberg proportions in cases. Am. J. Hum. Genet. *83*, 53–63.
2. Taylor, J., and Tibshirani, R. (2006). A tail strength measure for assessing the overall univariate significance in a dataset. Biostatistics *7*, 167–181.
3. Elston, R.C. (1991). On Fisher method of combining p-values. Biometrical J. *33*, 339–345.
4. Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S., et al. (2007). A genome-wide association study identifies novel risk loci for type 2 diabetes. Nature *445*, 881–885.
5. Conneely, K.N., and Boehnke, M. (2007). So many correlated tests, so little time! Rapid adjustment of *p* values for multiple correlated tests. Am. J. Hum. Genet. *81*, 1158–1168.
6. Zheng, G., and Ng, H.K.T. (2008). Genetic model selection in two-phase analysis for case-control association studies. Biostatistics *9*, 391–399.
7. Song, K., and Elston, R.C. (2006). A powerful method of combining measures of association and Hardy-Weinberg disequilibrium for fine-mapping in case-control studies. Stat. Med. *25*, 105–126.
8. Nielsen, D.M., Ehm, M.G., and Weir, B.S. (1998). Detecting marker-disease association by testing for Hardy-Weinberg disequilibrium at a marker locus. Am. J. Hum. Genet. *63*, 1531–1540.
9. Wittke-Thompson, J.K., Pluzhnikov, A., and Cox, N.J. (2005). Rational inferences about departure from Hardy-Weinberg equilibrium. Am. J. Hum. Genet. *76*, 967–986.
10. Wang, T., Zhu, X., and Elston, R.C. (2007). Improving power in contrasting linkage-disequilibrium patterns between cases and controls. Am. J. Hum. Genet. *80*, 911–920.
11. Freidlin, B., Zheng, G., Li, Z., and Gastwirth, J.L. (2002). Trend tests for case-control studies of genetic markers: power, sample size and robustness. Hum. Hered. *53*, 146–152.
12. Sasieni, P.D. (1997). From genotype to genes: doubling the sample size. Biometrics *53*, 1253–1261.
13. Zheng, G., Freidlin, B., Li, Z., and Gastwirth, J.L. (2003). Choice of scores in trend tests for case-control studies of candidate-gene associations. Biometrical J. *45*, 335–348.

DOI 10.1016/j.ajhg.2009.01.014.

Is the Tail-Strength Measure More Powerful in Tests of Genetic Association?

To the Editor: It is well known that Hardy-Weinberg equilibrium (HWE) is an important property in population genetics. Deviation from HWE among cases can provide evidence for a valid association.[1–4] Thus, it would be advisable to incorporate information from the HWE test for the improvement of power in detecting associated variants in genetic association studies. In the July 2008 issue of *The Journal*, Wang et al.[5] described a test statistic, the tail-strength (*TS*) measure,[6] for evaluation of the global null hypothesis, that the SNP was not associated with disease, which is a function of two p values: one from a logistic-regression test in a genetic association study and one from a HWE test in cases. The authors further extended the mean-based *TS* measure to a median-based measure (*TSM*) by measuring the deviation of each p value from its median value instead of its expected value. On the basis of simulation studies and real disease

Table 1. The p Value Correlation Coefficient between the HWE Exact Test and the Likelihood-Ratio Association Test

	Minor-Allele Frequency					
Genetic Model	0.05	0.1	0.2	0.3	0.4	0.5
Additive	−0.029	−0.0063	0.0014	0.014	0.012	−0.0004
Genotypic[a]	0.14	0.30	0.27	0.27	0.28	0.27
Dominant	−0.0037	0.024	0.041	0.063	0.10	0.12
Recessive	0.16	0.29	0.22	0.17	0.14	0.095

The results are based on 10,000 simulated data sets under the null hypothesis.

[a] The genotypic model means that the genotypes are coded as categorical variables and the major-allele homozygote is taken as a reference.

studies, the authors stated that the adopted *TS* measure was more powerful than the traditional logistic-regression test and that the type I error was also well controlled. However, we have two main concerns about these conclusions.

First, the two assumptions, which are required for deriving the exact distribution of *TS* and *TSM* statistics under the null hypothesis, hold only in certain scenarios. Violation of the two assumptions fails to obtain the exact distribution that the authors derived. The first assumption is that the two p values, of the HWE test and the logistic-regression test, are independent. This assumption may be violated, given that the two tests use the same case data. The authors mention that the adopted *TS* measure allows dependence between individual tests; yet, how to take the correlation into consideration was not discussed. The second assumption presumes the null distribution of HWE test p values to be uniform (0, 1), and this is breached when the exact test is applied for assessment of HWE.[7]

To evaluate the validity of these two assumptions, we generated 10,000 data sets of cases and controls (500 and 500) at various minor-allele frequencies (MAFs) (0.05, 0.1, 0.2, 0.3, 0.4, and 0.5) under the null hypothesis of no association between the SNP and disease status. We first assessed the correlations between the two p values for the HWE exact test and the association test (likelihood-ratio test) under four different genetic models, including additive, genotypic, dominant, and recessive models, respectively (Table 1). "Genotypic model" means that the genotypes are coded as categorical variables and the major-allele homozygote is taken as a reference. We observed nonignorable correlations between the two p values when a genotypic effect was modeled, whereas the correlations were low when an additive effect was modeled at different MAFs. For dominant and recessive models, the correlations between the two p values were not stable but were, instead, dependent on the MAFs. We also found that the HWE exact test p values did not correspond to a uniform distribution when we used the same simulated data, as described above (Figure 1). At the MAF of 0.05, the p values of the HWE tests were skewed to the left and did not follow a uniform

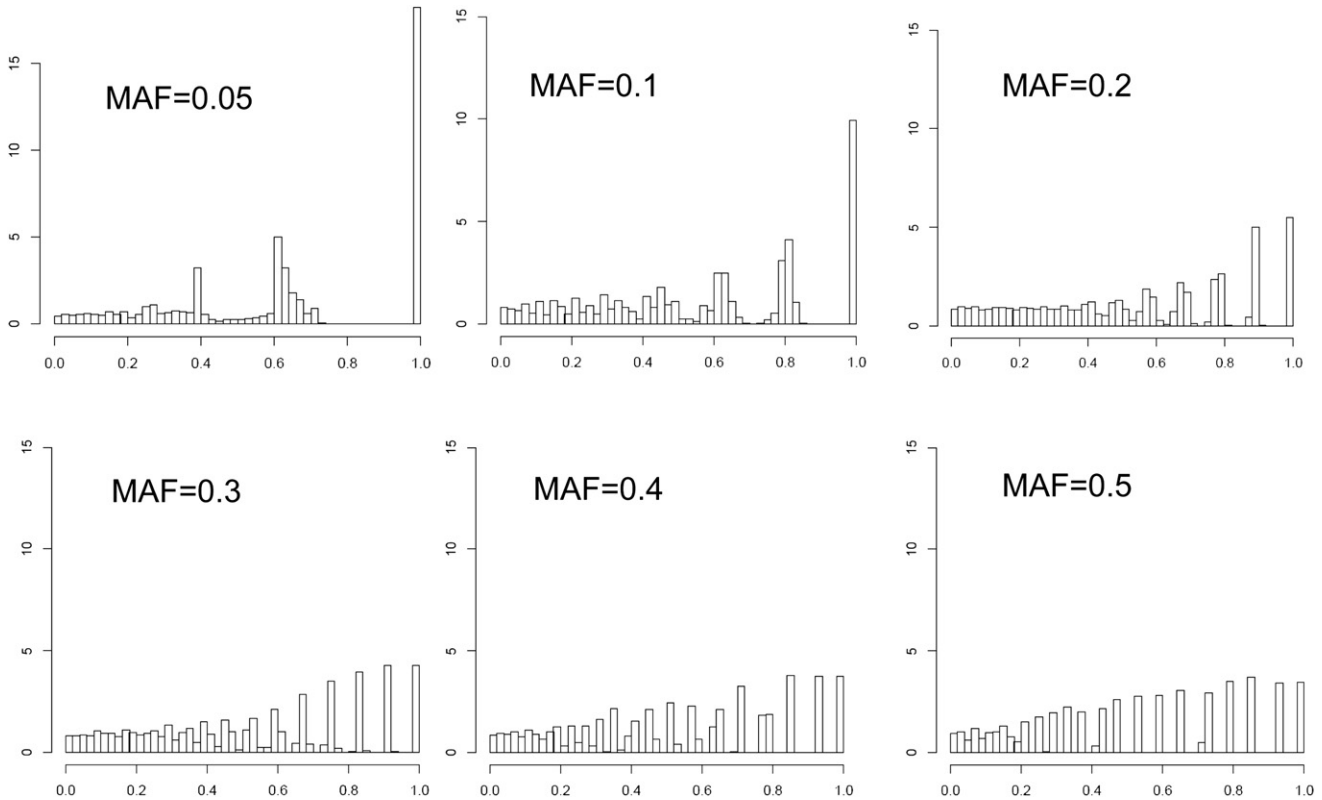

Figure 1. The Empirical Distribution of HWE Exact Test p Values under the Null Hypothesis at Various MAFs
Empirical distribution of HWE exact test p values in cases of 10,000 simulated data sets under the null hypothesis at various minor-allele frequencies (MAFs).

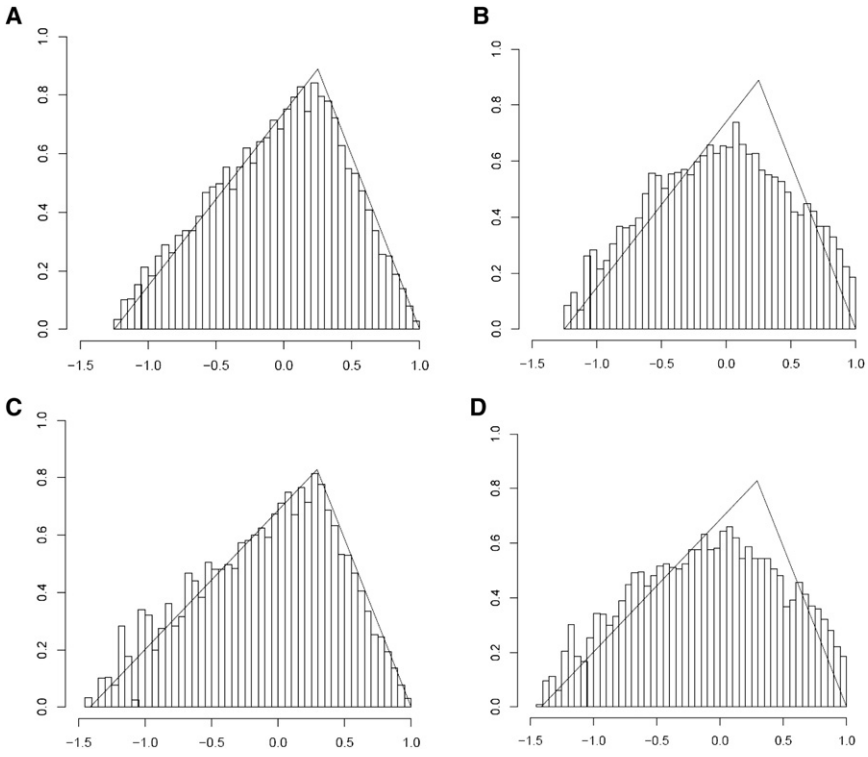

Figure 2. Comparisons of Exact and Empirical Distributions for *TS* and *TSM* Statistics under the Null Hypothesis

The exact and empirical distributions are represented by lines and bars, respectively. The results are based on 10,000 simulated data sets at a minor-allele frequency of 0.2. The exact p value is larger than the empirical p value, for both *TS* and *TSM*, at the tail of the distribution for a genotypic model.

(A) *TS*, additive model.
(B) *TS*, genotypic model.
(C) *TSM*, additive model.
(D) *TSM*, genotypic model.

distribution, and the degree of skew gradually diminished as the MAF increased. It is well known that under the null hypothesis, the p value based on a continuous test statistic has a uniform distribution over the interval [0,1], regardless of the sample size of the experiment.[8] What we observed in Figure 1 is due to the discreteness of the test. In essence, the HWE exact test is based on a discrete hypergeometric distribution of the data under the null hypothesis of HWE. Given the sample size and the MAF, only a finite number of possible distinct genotype configurations exists, and therefore, a finite number of possible p values generates a coarse distribution. When the sample size or MAF is small, the number of possible distinct genotype configurations will be small, with specific observed probabilities that deviate the distribution of p values from uniform. In particular, a spike close to a p value of 1 will appear for the most frequent sample configuration. As the sample size or MAF increases, more distinct genotype configurations increasingly resemble a uniform distribution.

To further evaluate the influence of these two assumptions on the analytical distribution of the *TS* and *TSM* statistics that the authors derived in the paper, we compared the analytical distribution of *TS* and *TSM* with the empirical distribution under different genetic models and MAF combinations, using the same simulated data sets. Results are shown in Figure S1 (available online). Specifically, for the additive model, when the MAF equals 0.05 or 0.1, the analytical p value is larger than the empirical p value at the right-hand tail of the distribution, which is more evident at a MAF of 0.05. However, the analytical and empirical distributions match pretty well at MAFs $\geq$ 0.2. Notably, when a genotypic model is used for calculation of the association test p value, the corresponding analytical and empirical distributions of the *TS* and *TSM* statistics do not fit well, regardless of the MAF, which can be attributed mainly to the nonignorable correlation between the two tests at various MAFs (violation of the independence assumption of the two tests). A similar result exists for a dominant or recessive genetic model when either the correlation between the two individual tests is nonignorable or the MAF is low (an MAF = 0.05 or 0.1). As an example, Figure 2 shows the empirical and the analytical distribution of the *TS* and *TSM* statistics at a MAF of 0.2 under the null hypothesis. We found that the analytical and the empirical distributions fit well for the additive model. However, it is obvious that the empirical and analytical distributions mismatch for a genotypic-effect model. The exact p value is less than empirical p value at the tails of the distribution for both *TS* and *TSM*. Therefore, we suggest the use of empirical distributions rather than exact distributions for *TS* and *TSM* measures in the practice of genetic association studies.

Second, we are concerned with the power of the *TS* and *TSM* statistics. To evaluate the empirical power, we performed another 10,000 simulations using the same parameters implemented in Wang et al.; i.e., simulation of data 1 in model 1 with $\beta_0 = -2$ and $\beta_1 = 0.3$. We calculated the power of the two marginal tests. Under the 0.05 significance level, the power for the HWE test is 0.053, which is only slightly larger than the type I error, and the power of additive effects in logistic regression is 0.59. Surprisingly, the empirical power of additive effects for the *TS* and *TSM* statistics, which combine the p values of the HWE exact test and the association test, is only 0.25 (0.27), which is lower than what the authors reported in their Table 4. We are unable to explain this discrepancy.

In summary, the *TS* statistic is useful for combining the information from HWE test and case-control association test to improve the power of detecting SNP effects in

genetic association studies. However, one needs to be cautious when using this statistic. On the basis of simulation results, we found that the analytical distributions of the *TS* and *TSM* statistics are influenced both by the MAF and by genetic models used in association tests. We suggest using the empirical p value, rather than the exact p value, in real situations. A more generalized statistic that does not depend on HWE-test significance in cases should be developed for the incorporation of HWE information and improvement of the power of genetic association studies.

Shizhong Han,[1,4] Li Ma,[2,4] Dawei Li,[1]
and Bao-Zhu Yang[1,3,*]
[1]Division of Human Genetics, Department of Psychiatry, School of Medicine, Yale University, New Haven, CT 06511, USA; [2]Department of Animal Science, University of Minnesota, St. Paul, MN 55108, USA; [3]Veteran's Administration Connecticut Healthcare Center, West Haven, CT 06516, USA
[4]These two authors contributed equally
*Correspondence: bao-zhu.yang@yale.edu

Supplemental Data

Supplemental Data include one figure and can be found with this article online at http://www.ajhg.org/.

Acknowledgments

This work was supported by the U.S. National Institutes of Health (career development award K01 DA024758 to B.Z.Y.). We are grateful for Joel Gelernter's valuable discussion and comments on this letter.

References

1. Feder, J.N., Gnirke, A., Thomas, W., Tsuchihashi, Z., Ruddy, D.A., Basava, A., Dormishian, F., Domingo, R. Jr., Ellis, M.C., Fullan, A., et al. (1996). A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat. Genet. *13*, 399–408.
2. Nielsen, D.M., Ehm, M.G., and Weir, B.S. (1998). Detecting marker-disease association by testing for HW disequilibrium at a marker locus. Am. J. Hum. Genet. *63*, 1531–1540.
3. Czika, W., and Weir, B.S. (2004). Properties of the multiallelic trend test. Biometrics *60*, 69–74.
4. Wittke-Thompson, J.K., Pluzhnikov, A., and Cox, N.J. (2005). Rational inferences about departures from HW equilibrium. Am. J. Hum. Genet. *76*, 967–986.
5. Wang, J., and Shete, S. (2008). A test for genetic association that incorporates information about deviation from Hardy-Weinberg proportions in cases. Am. J. Hum. Genet. *83*, 53–63.
6. Taylor, J., and Tibshirani, R. (2006). A tail strength measure for assessing the overall univariate significance in a dataset. Biostatistics *7*, 167–181.
7. Rohlfs, R.V., and Weir, B.S. (2008). Distributions of Hardy-Weinberg Equilibrium Test Statistics. Genetics *180*, 1609–1616.
8. Hung, H.M., O'Neill, R.T., Bauer, P., and Köhne, K. (1997). The behavior of the P-value when the alternative hypothesis is true. Biometrics *53*, 11–22.

DOI 10.1016/j.ajhg.2009.01.015.

Response to Zang et al. and Han et al.

To the Editor: In July 2008, we proposed a powerful test for the study of genetic association that incorporates information about deviation from Hardy-Weinberg proportions (HWP) in cases.[1] Two approaches were proposed: the mean-based tail-strength (*TS*) measure and the median-based tail-strength (*TSM*) measure. These measures combined p values from the likelihood ratio test (LRT) for association and the exact test for HWP. For both measures, we derived exact formulas to compute p values, and we also provided an approach for obtaining empirical p values with the use of a resampling procedure. The results showed a significant increase in power when using the proposed approaches. The type I errors were also well controlled with the additive model.

In their letter, Zang et al. report that when the underlying genetic model is not additive (recessive or dominant), there is a significant correlation between p values obtained from the LRT and the HWP test. Furthermore, they show that this correlation could lead to excessive false-positive probabilities if one uses the asymptotic formulas provided in our paper.

We agree that under certain situations the correlation between the two p values might not be ignored. However, in our original paper, we discussed limitations of the asymptotic null distributions of *TS* and *TSM*. We stated that "although the exact p values of *TS* and *TSM* are simple and straightforward to compute and interpret, the deviations of underlying assumptions might make the exact p values based on explicit formulas too conservative or too liberal." We therefore proposed an alternative approach for estimating empirical p values of *TS* and *TSM* with the use of a permutation procedure. For this permutation procedure, we resampled the SNP values by using the genotype frequencies calculated from the allele frequencies for both cases and controls. When the permutation procedure is applied, even if the assumptions underlying derivation of asymptotic null distribution are violated, one can still obtain accurate p values.

Tables 1 and 2 in Zang et al.'s letter show that the type I errors of the *TS* and *TSM* measures were inflated for the